



Predicting HLA CD4 Immunogenicity in Human Populations

Dhanda, Sandeep Kumar; Karosiene, Edita; Edwards, Lindy; Grifoni, Alba; Paul, Sinu; Andreatta, Massimo; Weiskopf, Daniela; Sidney, John; Nielsen, Morten; Peters, Bjoern

Total number of authors:
11

Published in:
Frontiers in Immunology

Link to article, DOI:
[10.3389/fimmu.2018.01369](https://doi.org/10.3389/fimmu.2018.01369)

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Dhanda, S. K., Karosiene, E., Edwards, L., Grifoni, A., Paul, S., Andreatta, M., Weiskopf, D., Sidney, J., Nielsen, M., Peters, B., & Sette, A. (2018). Predicting HLA CD4 Immunogenicity in Human Populations. *Frontiers in Immunology*, 9, [1369]. <https://doi.org/10.3389/fimmu.2018.01369>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



Predicting HLA CD4 Immunogenicity in Human Populations

Sandeep Kumar Dhanda^{1†}, Edita Karosiene^{1†}, Lindy Edwards¹, Alba Grifoni¹, Sinu Paul¹, Massimo Andreatta², Daniela Weiskopf¹, John Sidney¹, Morten Nielsen^{2,3}, Bjoern Peters^{1,4} and Alessandro Sette^{1,4*}

¹ Division of Vaccine Discovery, La Jolla Institute for Allergy and Immunology, La Jolla, CA, United States ² Instituto de Investigaciones Biotecnológicas, Universidad Nacional de San Martín, Buenos Aires, Argentina, ³ Department of Bio and Health Informatics, Technical University of Denmark, Kongens Lyngby, Denmark, ⁴ University of California San Diego, La Jolla, CA, United States

OPEN ACCESS

Edited by:

Clemencia Pinilla,
Torrey Pines Institute
for Molecular Studies,
United States

Reviewed by:

Karin Schillbach,
Universität Tübingen,
Germany
Silvia Deaglio,
Università degli
Studi di Torino, Italy
Lawrence J. Stern,
University of Massachusetts
Medical School,
United States

*Correspondence:

Alessandro Sette
alex@lji.org

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted
to T Cell Biology,
a section of the journal
Frontiers in Immunology

Received: 03 January 2018

Accepted: 01 June 2018

Published: 14 June 2018

Citation:

Dhanda SK, Karosiene E,
Edwards L, Grifoni A, Paul S,
Andreatta M, Weiskopf D,
Sidney J, Nielsen M, Peters B
and Sette A (2018) Predicting
HLA CD4 Immunogenicity in
Human Populations.
Front. Immunol. 9:1369.
doi: 10.3389/fimmu.2018.01369

Background: Prediction of T cell immunogenicity is a topic of considerable interest, both in terms of basic understanding of the mechanisms of T cells responses and in terms of practical applications. HLA binding affinity is often used to predict T cell epitopes, since HLA binding affinity is a key requisite for human T cell immunogenicity. However, immunogenicity at the population it is complicated by the high level of variability of HLA molecules, potential other factors beyond HLA as well as the frequent lack of HLA typing data. To overcome those issues, we explored an alternative approach to identify the common characteristics able to distinguish immunogenic peptides from non-recognized peptides.

Methods: Sets of dominant epitopes derived from peer-reviewed published papers were used in conjunction with negative peptides from the same experiments/donors to train neural networks and generate an “immunogenicity score.” We also compared the performance of the immunogenicity score with previously described method for immunogenicity prediction based on HLA class II binding at the population level.

Results: The immunogenicity score was validated on a series of independent data-sets derived from the published literature, representing 57 independent studies where immunogenicity in human populations was assessed by testing overlapping peptides spanning different antigens. Overall, these testing datasets corresponded to over 2,000 peptides and tested in over 1,600 different human donors. The 7-allele method prediction and the immunogenicity score were associated with similar performance [average area under the ROC curve (AUC) values of 0.703 and 0.702, respectively] while the combined methods reached an average AUC of 0.725. This increase in average AUC value is significant compared with the immunogenicity score ($p = 0.0135$) and a strong trend toward significance is observed when compared to the 7-allele method ($p = 0.0938$). The new immunogenicity score method is now freely available using CD4 T cell immunogenicity prediction tool on the Immune Epitope Database website (<http://tools.iedb.org/CD4episcore>).

Conclusion: The new immunogenicity score predicts CD4 T cell immunogenicity at the population level starting from protein sequences and with no need for HLA typing. Its efficacy has been validated in the context of different antigen sources, ethnicities, and disparate techniques for epitope identification.

Keywords: HLA, immunogenicity, immunodominance, epitopes, predictions, bioinformatics, TCR repertoire

INTRODUCTION

The identification of T cell epitopes has an important implication in several immunological contexts spanning from vaccine design to diagnostics in cancer, allergies, and infectious diseases fields. Most of the epitope identification is currently performed using bioinformatics prediction systems aimed to identify T cell immunogenicity and also to dissect the mechanisms underlying development of T cell responses. Currently, the majority of the T cell prediction methods are based on prediction of HLA binding affinity, which is a key requisite for human T cell immunogenicity. However, there is a lack of effective strategies able to predict immunogenicity at the population level, which is of particular importance when HLA typing data are not available. To overcome this issue, it is important to identify the common HLA binding affinity characteristics able to distinguish immunogenic peptides from non-recognized peptides. Two main classes of HLA molecules are important in the immunological context. Class I molecules presents epitopes to CD8 T cells, while class II molecules present epitopes to CD4 T cells. Prediction of HLA class I binding has reached high accuracy with area under the ROC curve (AUC) values greater than 0.9 (1–7), similarly HLA class II predictions have significantly improved in the most recent years reaching significant levels of accuracy (with AUC values in the range of 0.760–0.870) (8–10). However, HLA molecules are highly polymorphic and epitope prediction at the population level has to take into account this high level of heterogeneity.

We previously shown that in the case of HLA class I, focusing on 25–30 main HLA A and B allelic variants provides coverage of a large fraction of the general population (11). Similarly, in the case of HLA class II, about 40–50 allelic variants provide coverage of most frequent allelic variants (12). Prediction of HLA binding is usually performed with allele-specific algorithms, since binding motifs of different HLAs are rather diverse. However, in the case of HLA class II, it is also noted that a high degree of overlap exists between the epitope binding of different variants (13). Indeed, it was shown that the epitopes dominantly recognized are often capable of binding to many different HLA class II alleles. These epitopes (named promiscuous epitopes) account for 50% or more of the total responses at the population level (14).

The “7-allele method” was specifically optimized for prediction of HLA class II responses at the population level (15) based on the prediction of promiscuous epitopes. While this method is associated with significant predictive value, it is also expected that many of the peptides that are predicted or experimentally shown to bind HLA class II molecules may not induce T cell responses. This is because although HLA binding is necessary it is not sufficient by itself for T cell immunogenicity. Other factors such as antigen processing and the size of the TCR repertoire capable of recognizing any given MHC/epitope complex are key factors in ultimately determining immunogenicity (16–18).

In particular, it has been shown that the TCR repertoire is a key factor in shaping epitope immunodominance (19–23). In the case of HLA class I, different algorithms have been devised that evaluate a peptide sequence for the presence of certain amino acids, presumably interacting with TCRs, as a contributing factor to epitope's intrinsic immunogenic potential (24–26).

In the present study, we evaluate an approach to predict HLA class II immunogenicity at the population level, regardless of specific HLA haplotype, by training neural networks (NNs) with well-characterized sets of immunogenic epitopes dominant in general human populations. This approach could thus probe not only the influence of HLA binding but also potentially detect factors beyond HLA class II binding that would be encoded in the primary sequence of potential epitopes.

MATERIALS AND METHODS

Datasets

The datasets used for training were derived entirely from experimental data generated in our laboratory using congruent techniques as a mean to rely on tightly controlled datasets. In addition, we also utilized epitopes that were associated with positive tetramer data as part of the training, because tetramer data are regarded as “gold standard” of quality and specificity in analyzing T cell response. Conversely, the datasets used for validation were derived from scientific literature using a broad variety of techniques and antigens, and generated from different laboratories worldwide. This choice was made to ensure the robustness of the validation provided.

Training Dataset Assembly

We used 15-mer peptides derived from several datasets described in peer-reviewed articles or obtained by in-house studies following same experimental approach (Table 1). In some cases, the epitope sets were selected based on interim analysis and do not exactly match the final epitope lists in the published articles. The peptides were tested for immune recognition in cohorts of 5–150 donors by ELISPOT assays for one of the following cytokines: IFN γ , IL-5, IL-17, or IL-10. A full list of these epitopes is described in Table S1 in Supplementary Material. In total, 1,032 epitopes were selected as positives in this study. Negative peptides were selected from the same datasets listed in Table 1 following specific criteria: peptides should be negative in all tests, only peptides from proteins with at least one positive peptide recognized were included. In addition, any peptide tested more than once (due to several studies testing antigens/allergens from the same organism) giving opposite responses for the same donor was removed from the dataset. Overall, 5,739 negative peptides (Table S2 in Supplementary Material) were obtained. In some cases, set-specific adjustments in the criteria were necessary for technical reasons, as detailed below.

Mycobacterium Tuberculosis (TB) Antigens

We selected 65 previously known 15-mer epitopes identified from the vaccine candidate antigens and that captured 80% of the response (27–29).

Timothy Grass (TG) Known Allergens

Previous studies identified 20 epitopes that accounted for 79.5% of the total response to a set of TG-derived pollen antigens (Phl p allergens) in TG allergic individuals (14, 31, 32). Most of the datasets are composed by 15-mers as they were based on HLA class II binding prediction (15, 99). However, since some of those

TABLE 1 | Full list of datasets used in this study.**(A) Training datasets**

Antigen (s)	Peptide selection method	# of donors	Reference	# of epitopes	# of control peptides
<i>Mycobacterium tuberculosis</i>	Overlapping	18	(27)	65	53
	Predicted	28	(28)		1,043
	Overlapping	61	(29)		362
	Confirmed epitopes	61	(29)		137
Timothy grass	Overlapping	25	(14)	60	360
	Predicted	35	(30)		360
	Overlapping	21	(31)		6
	Overlapping	37	(32)		0
House dust mite (HDM)	Overlapping	20	(32)	52	6
Cockroach	Overlapping	19	(33)	71	521
Dengue antigens	Predicted	150	(34)	325	140
Erythropoietin	Overlapping	5	(35)	9	11
CRJ1 and CRJ2	Overlapping	54	(36)	30	18
Mouse allergens	Predicted	22	(37)	82	885
Novel HDM antigens	Predicted	20	(38)	105	186
Pertussis vaccine antigens	Overlapping	53	(39)	100	202
Ragweed allergens	Overlapping	25	(40)	15	183
Tetanus		20	(41)	28	98
ZIKA virus polyprotein	Overlapping	18	(Grifoni et al., unpublished)	48	529
Yellow fever virus polyprotein	Overlapping	42	(Weiskopf et al., unpublished)	42	639
Overall				1,032	5,739

(B) Validation dataset derived from literature

Antigen (s) (species)	# of donors	Reference	# of epitopes	# of control peptides
Acetylcholine receptor subunit alpha (<i>Homo sapiens</i>)	22	(42)	4	18
Circumsporozoite (CS) protein (<i>Plasmodium vivax</i> and <i>falciparum</i>)	22	(43)	4	4
Conserved hypothetical lipoprotein (<i>Francisella tularensis</i>)	10	(44)	3	10
Other protein (<i>Plasmodium falciparum</i>)	12	(45)	7	5
CS protein (<i>Plasmodium falciparum</i>)	64	(46)	7	10
CS protein (<i>Plasmodium falciparum</i>)	35	(47)	7	7
Api m 1 (<i>Apis mellifera</i>)	40	(48)	6	9
Myelin basic protein (<i>Homo sapiens</i>)	12	(49)	3	3
CS protein (<i>Plasmodium vivax</i>)	52	(50)	7	5
Acetylcholine receptor sub. γ and δ (<i>Homo sapiens</i>)	22	(51)	14	42
Acetylcholine receptor sub. α (<i>Homo sapiens</i>)	22	(52)	8	17
Glutamate decarboxylase 2 (<i>Homo sapiens</i>)	44	(53)	2	10
Structural polyprotein (<i>Rubella virus</i>)	10	(54)	4	7
Envelope glycoprotein D (<i>Human herpesvirus 1</i>)	24	(55)	6	6
Thyroglobulin and thyrotropin receptor (<i>Homo sapiens</i>)	15	(56)	5	10
Fusion glycoprotein F0 (<i>Morbillivirus</i>)	13	(57)	12	50
Poa p 5, <i>Poa pratensis</i> (<i>Kentucky bluegrass</i>)	13	(58)	9	8
Myelin basic protein (<i>Homo sapiens</i>)	20	(59)	6	7
Structural polyprotein (<i>Rubella virus</i>)	14	(60)	4	74
Acetylcholine receptor sub. δ and α (<i>Homo sapiens</i>)	58	(61)	12	33
Hev b 1 (<i>Hevea brasiliensis</i>)	19	(62)	2	2
Api m 1 (<i>Apis mellifera</i>)	10	(63)	7	6
TRAP (<i>Plasmodium falciparum</i>)	50	(64)	21	30
Nucleoprotein (<i>Morbillivirus</i>)	19	(65)	9	40
Genome polyprotein (<i>Hepatitis C virus</i>)	22	(66)	14	13
Subtilisin-like protease 6 (<i>Trichophyton rubrum</i>)	38	(67)	8	20
Blood groups Rh(D) and Rh(CE) polypeptides (<i>Homo sapiens</i>)	22	(68)	19	15
Myelin proteolipid and myelin basic protein (<i>Homo sapiens</i>)	16	(69)	7	14
Polyprotein Ent. virus B; Glut. Decarboxylase2 (<i>Homo sapiens</i>)	22	(70)	7	26
Gal d 1 (<i>Gallus gallus</i>)	14	(71)	2	1
Genome polyprotein (<i>Hepatitis C virus</i>)	10	(72)	5	122
Hev b 6 (<i>Hevea brasiliensis</i>)	16	(73)	4	12

(Continued)

TABLE 1 | Continued

(B) Validation dataset derived from literature

Antigen (s) (species)	# of donors	Reference	# of epitopes	# of control peptides
Bos d 9 (<i>Bos Taurus</i>)	10	(74)	2	5
Cha o 1 (<i>Chamaecyparis obtusa</i>)	19	(75)	10	24
Genome polyprotein (<i>Hepatitis C virus</i>)	22	(76)	12	257
Genome polyprotein (<i>Hepatitis C virus</i>)	41	(77)	18	33
Bos d 9, <i>Bos taurus</i> (<i>Bos Taurus</i>)	29	(78)	8	12
Cytochrome P450 2D6 (<i>Homo sapiens</i>)	80	(79)	28	29
Capsid protein VP1 (<i>Human parvovirus</i>)	19	(80)	8	54
Integrin beta-3 (<i>Homo sapiens</i>)	31	(81)	7	51
Genome polyprotein (<i>Hepatitis C virus</i>)	44	(82)	7	286
Equ c 1 (<i>Equus caballus</i>)	10	(83)	15	32
Merozoite surface protein 1 (<i>Plasmodium falciparum</i>)	48	(84)	10	18
Cry j 1 (<i>Cryptomeria japonica</i>)	12	(85)	4	33
Cha o 2 (<i>Chamaecyparis obtusa</i>)	19	(86)	6	36
Capsid protein VP1 (<i>Adeno-associated virus</i>)	16	(87)	28	62
Non-specific lipid-transfer protein (<i>Prunus persica</i>)	15	(88)	3	5
Aquaporin-4 (<i>Homo sapiens</i>)	32	(89)	6	10
UniProt:B8ZU53 (<i>Mycobacterium leprae</i>)	152	(90)	8	1
Pas n 1 allergen (<i>Paspalum notatum</i>)	18	(91)	4	11
Pen a 1 allergen (<i>Farfantepenaeus aztecus</i>)	16	(92)	15	13
Genome polyprotein (<i>Tick-borne encephalitis virus</i>)	47	(93)	26	46
Other wolf or dog protein (<i>Canis lupus</i>)	25	(94)	18	12
Can f 5 (<i>Canis lupus</i>)	24	(95)	25	31
Botulinum neurotoxin type A (<i>Clostridium botulinum</i>)	25	(96)	6	13
Genome polyprotein (<i>Rhinovirus A and C</i>)	20	(97)	15	34
Botulinum neurotoxin type A (<i>Clostridium botulinum</i>)	14	(98)	6	14
Overall			530	1,758

epitopes were not 15-mers, to compare those with the rest of the dataset longer epitopes were dissected into the composing 15-mers and each 15-mers belonging to the longer peptides has been classified as a positive, with the same process being used for negative peptides. In addition, 19 peptides were described to cover an NTGAp19 peptide pool, which were selected to encompass at least 40% of the total IL-5 response directed against all NTGA peptides screened (30).

House Dust Mite (HDM) Allergens

The peptide set included the 34 most dominant epitopes cumulatively accounting for 90% of the total allergen-specific response detected in our screen (32). Analogous to the TG set, longer regions were deconstructed into 15-mers, which yielded 52 peptides in total.

Cockroach (CR) Allergens

71 most dominant epitopes were selected based on total spot forming cells (SFC) values greater than 1,000 (33).

Dengue (DENV) Antigens

Peptides predicted to bind various frequent DRB1 alleles were tested in about 10 HLA-matched donors. The sets comprised 325 epitopes, positive in at least two donors with PBMC derived from normal blood donors from the Colombo (Sri Lanka) region that were seropositive for DENV antibodies and thus representative of natural infection (34). Negative peptides were those tested in at least 10 donors and found to be uniformly negative.

Erythropoietin

Tangri et al. screened overlapping peptides and reported nine epitopes recognized by at least 40% donors (35).

CRJ1 and CRJ2 Japanese Cedar Allergens

This set contained overlapping 15-mers spanning the CRJ1 and CRJ2 allergens (36). We selected 30 dominant epitopes based on average response magnitude of >100 SFC (sum of IL-5 and IFN γ) in either of two group of allergic donors: those who lived in Japan for extended periods of time and USA sensitized donors who had not lived in Japan. A total of 18 control negative peptides were derived from allergens CRJ1 and CRJ2 and selected based on a response frequency of one donor or less and an individual SFC response <100 SFC.

Mouse Allergens

Peptides derived from mouse allergens, largely selected by the 7-allele algorithm were tested in 22 donors (37). A total of 89 dominant epitopes were defined on the basis of total SFC >150 and recognized in at least two donors.

Novel House Dust Mite Antigens

The peptides screened were predicted with the 7-allele method from 96 HDM (novel and known) proteins in 20 HDM allergic donors (38). We selected the 106 more dominant epitopes, recognized in multiple donors and with an overall magnitude of >300 SFC total (accounting for about 50% of the total response).

Pertussis Vaccine Antigens

The peptide set was comprised of 16-mers overlapping by eight residues, spanning the entire sequence of the antigens. We selected the top 100 epitopes recognized in at least 4 of the 53 total donors analyzed, and accounting for approximately 75% of the total response (39).

Ragweed Allergens

This set included 16-mers overlapping by eight amino acids, spanning the entire sequence of the antigens (40). A total of 15 epitopes accounting for 75% of the total response was selected. If variants were present, the most common variant was selected.

Tetanus Toxoid (TT) Antigen

We selected a set of 28 epitopes, recognized in at least 2 out of 20 donors tested (41), and predicted by the 7-alleles method (15). As a control, we selected a set of 57 peptides, which were studied but not recognized, neither in the Immune Epitope Database (IEDB, www.iedb.org) nor in the study by Antunes et al. (41), and an additional set of 41 peptides that were not predicted by the method and also neither recognized in the study by Antunes et al. nor identified in the IEDB as positive human responses. In the case of the third set of 41 peptides, there were 261 15-mers in the Tetanus set. Among them 124 were predicted to be binders with predicted 7-allele median percentile rank ≤ 20.0 . Out of the 137 non-predicted peptides, those with predicted 7-allele median percentile rank > 40.0 were selected (67 peptides) for screening to be included in non-predicted AND non-epitope set. From this list, we eliminated peptides that were overlapping by more than five AA residues with any of the epitope (recognized in our study or annotated as positive in IEDB). The remaining 41 peptides were included in the set of “control peptides” that were not predicted and neither recognized in the Antunes et al. study nor identified as positive response in IEDB.

ZIKA Virus (ZIKV) Antigens

A set of 15-mer peptides spanning the entire sequence of the ZIKV proteome was tested with a 14 days re-stimulation protocol in 18 donors. A total of 48 epitopes were defined as being positives in at least two donors (Grifoni et al., unpublished).

Yellow Fever (YF) Antigens

The set of epitopes tested includes 94 previously described YF CD4⁺ T cell epitopes with known HLA class II restriction (IEDB) and sets of peptides predicted to bind different HLA DRB1 molecules. CD4⁺ T cells from 42 donors vaccinated with YF17 vaccine were co-cultured with autologous antigen-presenting cells and HLA-matched YF DRB1 predicted peptides. After 14 days, IFN γ response against individual peptides was determined as previously described (100). Epitopes were defined as peptides eliciting an SFC of 664 SFC/10⁶ or more. This resulted in the identification of 42 unique peptides (Weiskopf et al., unpublished).

IEDB Validation Datasets

To generate additional datasets to evaluate the performance of the various predictive schema, we sought to identify literature records reporting overlapping peptide studies. Accordingly, we

queried the IEDB for papers which contained both positive and negatives records curated in the paper, related to HLA class II restricted T cells. This query identified 870 papers; which were further refined by filtering by “overlapping” mentioned in the abstract, resulting in 183 records.

The abstracts of those records were manually inspected, to select papers truly related to study of immunogenicity of overlapping peptide sets. At this stage, we excluded records relating to Phl p, TT, TB (already represented in the previous sets) and studies based on transgenic mice to obtain 102 relevant papers.

We next removed papers where the peptide size was less than 15, or where less than 10 donors were studied (resulting in 82 papers). Each of these 82 papers was manually inspected and additional papers were discarded upon manual inspection for a variety of reasons, including the paper not reporting testing for full sets of overlapping peptides, ambiguous reporting of negative results or peptide size tested, no clear discrimination between positive and negative responses, testing pools of peptides with no deconvolution, and similar problems.

This resulted in a final selection of 57 papers (Table 1). For each paper, based on the data disclosed and on the author's interpretations, we captured the most dominant epitopes accounting for the majority of responses and/or consistently positive in multiple donors. We selected peptides that were consistently negatives as corresponding negative controls. In studies where large numbers of donors were tested and essentially all peptides were positives, we selected the peptides positives in one or more donors. A list of PUBMED IDs, and the criteria used to select the “top” epitopes and the “bottom” negative controls is provided in Table S3A in Supplementary Material. A list of positive and negative control peptides is provided in Table S3B in Supplementary Material.

Tetramer Training Dataset

A dataset corresponding to epitopes described as positive in tetramer staining experiments was downloaded from IEDB (accessed June 2015) (101) using the following selection criteria: “Positive Assays Only, Epitope Structure: Linear Sequence, T Cell Assays: qualitative binding/multimer/tetramer (tetramer), No B cell assays, No MHC ligand assays, MHC Restriction Type: Class II, Host Organism: *Homo sapiens* (human) (ID:9606, human).” The exported dataset was filtered keeping only 15-mer epitopes for which a source antigen protein ID was available. For each unique positive peptide, we took its source protein sequence using the antigen genome ID and scanned that protein for all possible 15-mers overlapping by 10 amino acids. The original positive peptide was then considered as an immunogenic one and the rest of the obtained peptides were used as negatives. The tetramer dataset had 124 unique positives and 5,319 negatives that are presented in Table S4 in Supplementary Material.

Artificial Neural Network (ANN)-Based Predictions Using *NNAlign* Method

The NN training for peptide sequences was performed using the *NNAlign* method (102). The method uses classified peptide data for training and identifies nested shorter sequence patterns that constitute an informative motif to separate positive from negative examples. As an input to *NNAlign*, we used sequences of our

15-mer peptides and their assigned observed immunogenicity score (1.0 for immunogenic and 0.0 for non-immunogenic). The method was trained using extensive cross-validation where part of the data is left out of the training process and is used for evaluation purpose only. For each peptide, the method returns a predicted score between 0.0 and 1.0, with high values identifying more immunogenic peptides and low values non-immunogenic peptides. The *NNAlign-1.4* software package was downloaded from <http://www.cbs.dtu.dk/services/NNAlign/>. The method was trained for each possible motif length varying from 1 to 15. The data for cross-validation was split based on common motifs within peptides with a maximum overlap to nine and varying the motif length. Input peptides were encoded using Sparse and BLOSUM schemes. No rescaling was done to the input data. We also chose to preserve repeated flanks in the original data and do not realign networks with offset. The method was trained with 5 hidden neurons using 10 seeds for each network architecture. It is possible that other encoding approaches, choice of NN design, or choice of other learning algorithms could have led to better results, but such a comparison was outside the scope of our current manuscript.

Receiver Operating Characteristic (ROC) Curves and AUC Values

To measure how different approaches are capable of classifying peptides into epitopes and non-epitopes, ROC curves were used (103). Varying cutoff for predicted scores, peptides were classified into immunogenic and non-immunogenic and the numbers of true positives (TPs) and false positives (FPs) were obtained. The ROC curve was made by plotting TP rate as a function of FP rate at each cutoff. AUC is a useful measure for assessing predictive performance of a prediction method. AUC values range from 0.5 to 1, where 0.5 corresponds to random and 1 to perfect predictions. The AUC value can be interpreted as the probability that the predicted score for a randomly chosen immunogenic peptide is higher than the score of a randomly chosen non-immunogenic peptide.

HLA Binding Predictions

We utilized the previously described 7-allele method (15) to derive HLA binding propensities. The 7-allele method predicts immunogenicity based on the median percentile predicted binding of seven alleles representative of the binding motifs most commonly recognized in the general human population, and is available on the IEDB website (104).

Generation of Two-Sample Logo

The two-sample logo was created with 15-mer peptides (15 residues from N-terminal were extracted in case of longer peptides) from all the datasets combined, for epitopes and non-epitopes. For two-sample logo, both epitopes and non-epitopes datasets (in FASTA formatted files) were submitted to the online tool (<http://www.twosamplelogo.org/cgi-bin/tsl/tsl.cgi>) with default settings except for *p*-value, which was set to 0.01 and resolution of 600 dpi (105).

Statistical Analysis

The statistical analysis was performed using Prism 7 (Graph-Pad Software, San Diego, CA, USA). The non-parametric Wilcoxon

matched-pair signed rank test with method of Pratt was utilized to assess the significance differences between sets of different AUC values.

RESULTS

Derivation and Validation of an ANNs-Derived Immunogenicity Score

We assembled T cell epitope datasets from different previously published peptide screening studies performed in our laboratory (Table 1). In all cases, peptides were screened using ELISPOT assays to detect which peptides stimulated secretion of cytokines. Table 1 summarizes the number of donors that were screened for each peptide set and if the peptides were selected to overlap specific antigens, or if they were selected based on predicted binding affinity. Dominant epitopes accounting for a majority of the T cell responses as described in more detail in the Section “Materials and Methods” were considered positives (Table S1 in Supplementary Material, *N* = 1,032 peptides). Peptides that did not give any response in any donor but that came from proteins for which at least one peptide was positive, were considered negatives (Table S2 in Supplementary Material, *N* = 5,739 peptides). This additional criterion for negative classification was used to ensure that the lack of recognition was not simply due to lack of availability of the source protein necessary for antigen presentation.

This initial dataset was used to train an ANN-based method called *NNAlign* (102). The *NNAlign* method takes an unaligned peptide set and aims to find a linear sequence core within the peptides, which differentiates the positive (immunogenic) from the negative (non-immunogenic) peptides. The length of the sequence core was varied systematically from a single residue to 15 residues immunogenicity score (ranging 0–1) for each variation is retrieved, and prediction quality was assessed using fivefold cross-validation. Several sequence core lengths showed AUC values greater than 0.7 which is generally considered as a good prediction quality value, suggesting that the ANNs shows differences between positive and negative peptides based on the peptide motif. In terms of sequence motif length, the cross-validation did not indicate a clear optimal length, as the prediction performance was similar for motif lengths between 3 and 12 (Figure 1). A motif length of nine residues is consistent with the known size of peptide core region engaging HLA and TCR. For this reason, a motif length of nine was selected for the following analyses.

Combining Immunogenicity and HLA Binding Predictions

To consider both HLA binding and the immunogenicity prediction (which presumably incorporate the capacity of being recognized by TCR), we combined our ANN-based immunogenicity predictions with HLA class II binding predictions. Only one method has been described to predict epitopes based on HLA binding at the population level, namely the 7-allele method, which was previously empirically optimized based on immunogenicity datasets (15).

To combine immunogenicity and HLA binding scores, we used the median percentile rank score (*HLA_score*) of the

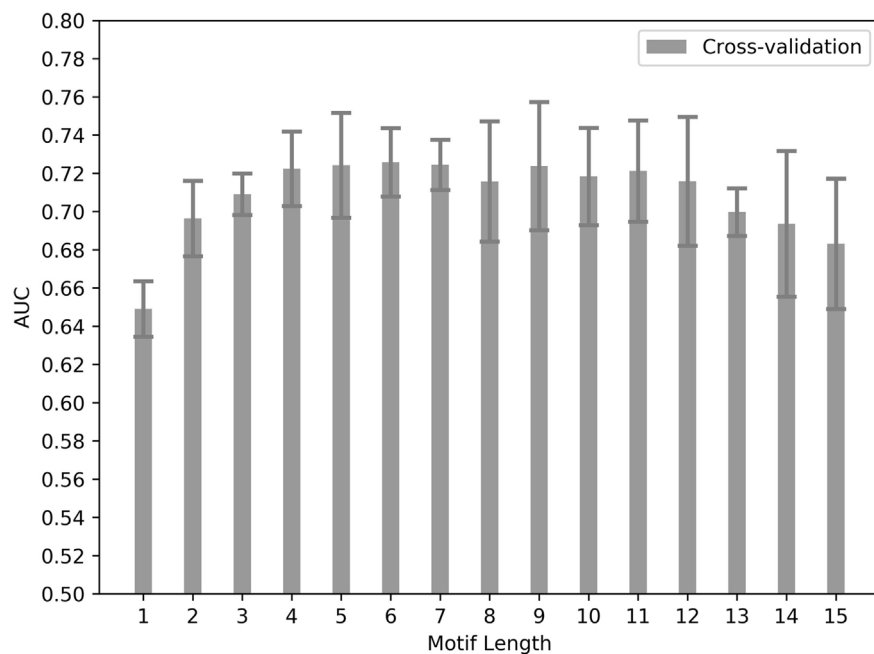


FIGURE 1 | Predictive performances for different motif lengths. Bars show cross-validation performance for the training dataset. Area under the ROC curve (AUC) values are shown for each artificial neural network training done by choosing different sequence lengths to define a preferred sequence motif within a 15-mer peptide. Error bars show SD of the five cross-validation sets.

7-allele method (ranging from 0 to 100) and combined it with our NN-based immunogenicity score after converting it to a percentile score, so that it would also range from 0 to 100 and could be comparable with the HLA_score, using the formula (Imm_score) = $(1 - \text{neural network immunogenicity}) \times 100$. The two scores were combined as follows:

$$\text{Combined score} = \alpha \times \text{Imm score} + (1 - \alpha) \times \text{HLA score. (1)}$$

Next, we systematically varied the value of α in the interval of $0 \leq \alpha \leq 1$. From the equation above, when $\alpha = 1$ the results depend only on the immunogenicity predictions by the NN, while with $\alpha = 0$ only HLA binding predictions are used to define immunogenicity.

To assess the performance of the immunogenicity score, the 7-allele method and their combination, we used independent literature-based datasets. Specifically, we searched the IEDB for papers which described results of testing overlapping peptide sets related to human HLA class II restricted T cells. These epitope sets thus represent a broad range of studies, representing a “real-life” portrait of epitope identification studies performed in the worldwide scientific community. These epitope sets are listed in Table S3A in Supplementary Material and described in more detail in the Section “Materials and Methods” and the sequences are provided in Table S3B in Supplementary Material. Overall, a total of 57 different sets derived from independent literature studies were curated, entailing a total of 530 positive and 1,758 negative peptides. **Figure 2** depicts the predictive performance of the combined score, displaying the average of the different AUC values obtained for each of the different datasets. The 7-allele method

was associated with AUC values of 0.695, and the immunogenicity score was associated with an average AUC value of 0.670. In terms of combination of the two algorithms, the performance increased and reached a peak at 0.71 for an α value of 0.50.

Performance of the Immunogenicity Score, Eliminating Redundancy Between Training and Testing Datasets

It is expected that inclusion of additional data points would increase the performance of an NN model. Accordingly, we incorporated an additional dataset of CD4 T cell epitopes identified by tetramer mapping studies. We reasoned that this would provide high quality epitopes since the tetramer-staining assay is commonly considered a “gold standard” assay for epitope characterization. The dataset was obtained by querying the IEDB for 15-mer peptides that were tested positive by tetramer staining assays. For each positive peptide, its source protein was scanned for 15-mer peptides overlapping by 10 amino acids, with the positive peptide sequences being removed and the remaining peptides used to construct a negative dataset. The final tetramer dataset is composed of 124 unique positives and 5,319 unique negative peptides (Table S4 in Supplementary Material).

The datasets utilized to train and evaluate the NN models contained some redundancies, which could affect the evaluation and inflate performance. To avoid this issue, we eliminated any redundancy between the training set (**Table 1** and tetramer set combined) and the validation set of the 57 independent studies (**Table 1**) by filtering out any peptide sharing a common 9-mer sequence.

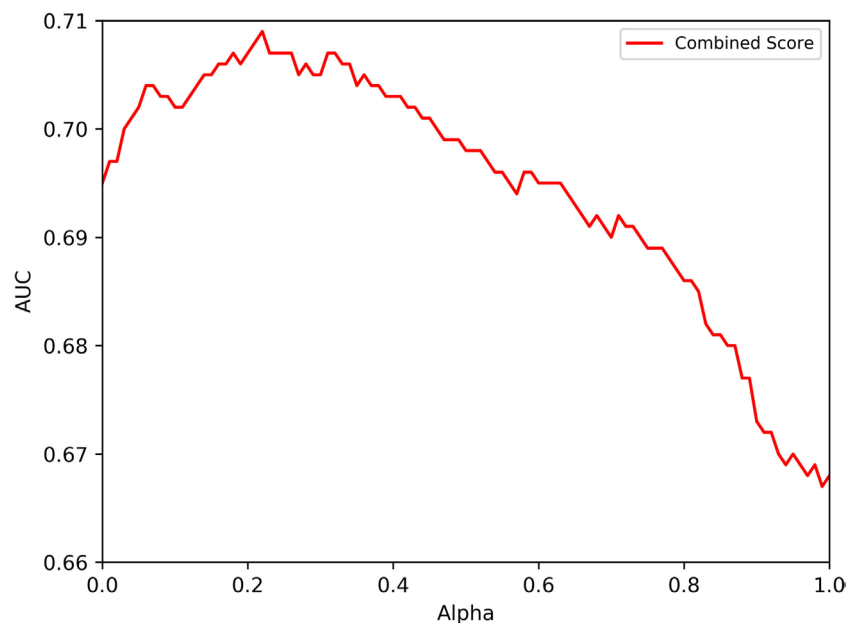


FIGURE 2 | Predictive performances obtained combining HLA binding and immunogenicity scores. The figure shows the performance dependency on an α coefficient used to combine HLA binding and immunogenicity scores. The model trained on the training dataset described in the text and validated on independent literature datasets, also described in the text.

In the analysis performed, a clear optimal alpha was not observed. The data in **Figure 2** seemed to indicate an optimal alpha around 0.2–0.3, while the analysis from **Figure 3** indicates two optimal peaks at about 0.4 and 0.6. Since the data in **Figure 3** are inherently more reliable because of training with a higher number of data points, we empirically selected 0.4 as the alpha to include in the next set of analyses. When this analysis was performed, the 7-allele method prediction and the immunogenicity score were associated with similar performance (average AUC values of 0.703 and 0.702, respectively) while the combined methods again afforded gain in performance, reaching an average AUC of 0.725 (**Figure 3**). This increase in average AUC values of the combined methods is significant when compared with the average AUC values of the immunogenicity method with a p value of 0.0135 using Wilcoxon matched-pairs signed rank test, and a strong trend toward significance when compared to the 7-allele method with a p value of 0.0938. These results, together with the ones obtained with the tetramer dataset confirm that both the 7-allele and the immunogenicity score method had significant predictive value on their own which are in both cases enhanced by their combination.

Two-Sample Logo of a General Immunogenicity Motif

Next, we analyzed the epitopes and non-epitopes from all the datasets combined for their positional residue conservation and plotted two-sample logo using 15 residues from the N-terminus of the peptides (105) (**Figure 4**). The two-sample logo represents amino acids which were significantly different in epitopes and non-epitopes based on p -value (<0.01) calculated using t -test. Amino acid residues enriched in the epitope dataset are mostly

positively charged, while amino acid residues depleted in the epitope dataset (and enriched in the non-epitope dataset) are mostly negatively charged. In other words, epitopes have higher numbers of positive charged residues like arginine (R) or lysine (K) at positions 9th and 11th–14th, whereas non-epitopes contained aspartate (D) and glutamate (E) at positions 7th–9th and 11th–13th. A preference for hydrophobic residues is also observed [such as proline (P) and alanine (A)] in non-epitopes, whereas isoleucine (I), phenylalanine (F), and asparagine (N) are enriched in the epitope set. To further address the significance of the logo, we split the dataset into five sets, where each set contains 80% of the total dataset, the results in Figure S1 in Supplementary Material confirm that the most prevalent feature revealed by the logo are in consistent with the two-sample logo created using whole dataset (**Figure 4**). These results suggest that some of these preferences may be contributing to T cell recognition or MHC binding or represent a result of processing enzymes. These possibilities will be addressed in future studies.

Epitope Prediction Threshold and Implementation of an Online Tool

We next determined the performance of the combined score using different cutoff values ranging from 0 to 100 (**Table 2**) for each study. To this end, we calculated the performance of overlapping datasets derived from literature at different threshold settings using the percentile combined score at $\alpha = 0.4$. As a first step, for each study we calculated the numbers of: true negative (TN) defined as non-immunogenic peptides predicted as non-immunogenic, FP defined as non-immunogenic peptides predicted as immunogenic, false negative (FN) defined as immunogenic

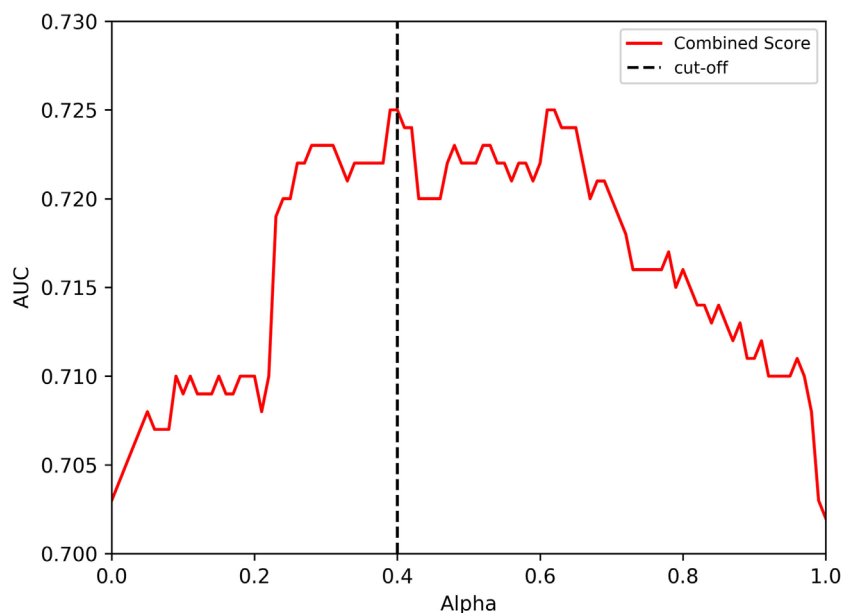


FIGURE 3 | Performance of independent literature datasets with combined approach and varying degree of alpha on the model trained with initial, in-house and tetramer datasets. The prediction values from HLA score and immunogenicity score using different values of alpha are shown. A cutoff of 0.4 value for alpha is also highlighted by a dotted line.

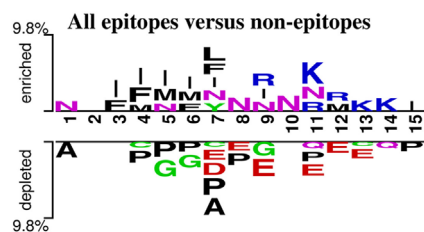


FIGURE 4 | Two-sample logo created using epitopes and non-epitopes in all the data (p -value < 0.01). The immunogenicity motifs for epitopes and non-epitopes were derived from the combination of all the datasets.

TABLE 2 | Performance of overlapping dataset derived from literature at different threshold settings using the percentile combined score.

Threshold	Average sensitivity	Average specificity	Total peptide to be synthesized (average)
8	20	91	13
18	31	85	21
36	51	65	39
43	59	59	46
66	75	37	68

The performance for each study is calculated individually and then averaged.

Sensitivity = $(TP/TP + FN) \times 100$, specificity = $(TN/TN + FP) \times 100$, % total peptide to be synthesized = $[(TP + FP)/(TP + TN + FP + FN)] \times 100$.

TN, true negative; FP, false positive; FN, false negative; TP, true positive.

peptides predicted as non-immunogenic, and TP defined as the immunogenic peptides predicted as immunogenic. Based on these values we calculated sensitivity $[= (TP/TP + FN) \times 100]$ and specificity $[= (TN/TN + FP) \times 100]$. Finally, we determined that

cutoff values of 8, 36, and 66 allowed capturing, respectively, 20, 51, and 75% of the epitopes with a corresponding specificity of 91, 65, and 37%. We also estimated the fraction of peptides needed to test in order to observe a defined fraction of epitopes using the following formula: $[(TP + FP)/(TP + TN + FP + FN)] \times 100$. A value of 43 was associated with equal sensitivity and specificity (59). To make this approach user friendly, we also implemented an online version of this algorithm (**Figure 5**). The tool is freely available in the IEDB website at <http://tools.iedb.org/CD4episcore/>.

DISCUSSION

Bioinformatics predictions to identify T cell epitopes are frequently used in the context of designing and testing vaccines and diagnostics for infectious diseases, allergies, and cancer. While several HLA allele-specific predictive algorithms (10) and T cell epitopes predictive strategies based on MHC class II binding have been described (106–108), development of effective strategies to predict immunogenicity at the population level are lacking and remain therefore of significant interest. This is important, since in the real-life applications most often encountered HLA typing data is often unavailable.

Here, we report an approach to identify sequence motifs distinguishing immunogenic peptides recognized by CD4+ T cells from non-recognized peptides, independent of the restricting HLA class II allele. We confirm that the previously described 7-allele method (15) is effective in predicting epitopes and could narrow the range of peptides to be used for biological testing. Importantly, we find significant improvements of a combined HLA binding + immunogenicity approach over immunogenicity predictions alone and a strong trend toward significance of a

The screenshot displays the IEDB Analysis Resource website. At the top, there is a blue header with the text "IEDB Analysis Resource". Below this is a navigation bar with links: Home, Help, Example, Reference, and Contact. The main heading is "CD4 T cell immunogenicity prediction". The interface is divided into several sections:

- Specify Sequence(s)**: This section contains a large text input field for "Enter epitope sequence(s) in PLAIN or FASTA format". Below this is a button labeled "Browse..." and the text "No file selected.".
- Choose a prediction method**: This section has a label "Prediction method:" followed by a dropdown menu currently set to "IEDB recommended".
- Specify Output**: This section contains three input fields:
 - "Sort Peptides by:" with a dropdown menu set to "Position in Protein".
 - "Select maximum percentile rank threshold:" with a dropdown menu set to "20".
 - "Email address (optional)" with a text input field.

At the bottom right of the form, there are two buttons: "Submit" and "Reset".

FIGURE 5 | Screenshot for home page of immunogenicity prediction server.

combined HLA binding + immunogenicity approach over HLA binding predictions alone.

The machine learning algorithm we applied (*NNalign*) was developed to identify sequence motifs of a specific length that distinguish peptide sets—in our case immunogenic from non-immunogenic peptides. We found that motif lengths between 8 and 11 residues gave the best performance in the classification of the different datasets. This motif length is in line with what has been described with epitope residues in contact with the T cell receptor, and the length of the epitope core binding characteristic of the HLA class II molecule, which is also about nine amino acids long (20, 109).

The fact that the increase over predictions performed on HLA binding alone is rather small suggests, in line with previous studies, that HLA binding is a dominant force in shaping the repertoire of T cell epitopes. It is also possible, however, that this relatively small increase might be related to coordinate evolution between HLA binding and antigen processing and TCR recognition as suggested before by other studies (110).

Since the method was derived on immunogenicity outcomes only, it is possible that the motif defined herein is not only related to HLA binding but also incorporates overall preferences for TCR residue contacts. However, given the unbiased nature in which it was derived, it cannot be ruled out that the method may also reflect completely different processes, such as modulation by

HLA-DM or increase in HLA binding stability over affinity is the actual source of the motif (111).

The predictive ability of very short motifs (3, 4 residues) is striking. Potential structural or mechanistic bases for this could be reflective of dominant influence of short stretches of residues incorporating dominant residues for HLA binding in close proximity to residues also dominant in TCR recognition (15). Examining the residues in the motif suggests that peptides with small amino acid side chains are avoided in the middle of the motif, while residues with longer side chains are overrepresented. This is qualitatively similar to what we had previously found for HLA class I restricted epitopes, and which has been reported in experimental studies using single residue substitutions (112, 113). This further supports that the motif identified coincides with properties of peptides more likely to engage a TCR. The F, M, L enrichment in the positions close to the N-terminus maybe at least in part corresponding to the P1 anchor of the MHC-II, which has similar specificity in several loci and allelic variants.

Our results have been trained over an extended set of data, derived from different methodologies and from populations of diverse ethnicities, and related to infectious diseases, allergy, and autoimmunity. The tetramer-trained algorithm seems to perform better, despite a bias toward certain HLA alleles and possible inclusion of many epitopes in negative set (i.e., other epitopes from the same protein other than the tetramer considered). We

speculate that this may be due to the fact that tetramer epitopes represent usually dominant epitopes which in turn have been shown to correspond to promiscuous HLA binders. Overall, the combined training sets corresponded to over 14 thousand peptides, from over 300 different antigens and tested in over 2,500 different human donors. We believe this is an important aspect of our study, as it ensures that our building model (as related to both the 7-allele method, the immunogenicity score and the combined approach) are valid irrespective of antigen source, different ethnicities and disparate techniques for epitope identification. Our prediction method may be useful for generating off the shelf vaccine peptide libraries for pathogens or common tumor markers. Conversely, this method may be useful for an optimum selection of peptides covering individualized tumor derived neo-epitopes after NGS sequencing in HLA-typed individuals.

The algorithm is available on the IEDB website (101), and we estimate that the use of the combined immunogenicity score and 7-allele method will allow capturing 50% of the total epitopes by synthesis of 24% of the total possible overlapping 15-mers. This would translate in coverage of a 300 residues protein with 72 15-mer peptides. Future improvements of T cell epitope predictions may benefit from the increased availability of large scale datasets of peptides eluted from HLA class II molecules, datasets of specific TCRs recognizing epitopes, and datasets unraveling the role of mediators in the MHC class II processing pathway such as HLA-DM.

Even with this approach the AUC values are lower than for MHC-I analysis (1). However, it should be kept in mind that these AUC values refers to prediction at the population level encompassing T cell with diverse restriction, while the higher AUC values for MHC-I usually refers to allele-specific predictions. However, the application of the current approach from MHC-II to MHC-I, faces specific challenges. In MHC-I it is thought there is much more HLA-specific selection of epitopes, arguing against a straightforward application of the current approach, but it is possible that the alpha analysis could identify any HLA-independent components. Finally, it will be of interest to develop a similar approach to develop HLA agnostic predictors of HLA class I epitopes. Recent data suggest that it is possible to empirically develop HLA class I epitope “megapools” that

afford coverage of general populations, irrespective of ethnicity (114, 115). Future studies will be focused on similar methods for HLA-agnostic prediction of class I restricted epitopes.

ETHICS STATEMENT

Human data have been previously published and extracted from IEDB database (www.IEDB.org).

AUTHOR CONTRIBUTIONS

SD, EK, LE, SP, MA, and JS compiled and analyzed the data. SD, EK, AS, and BP wrote and edited the manuscript. AG and DW contributed the data. AS, MN, and BP conceived and supervised the project.

FUNDING

This work has been supported by the following grant(s) of National Institute of Allergy and Infectious Diseases: 10.13039/100000060 HHSN272201200010C, HHSN272200900042C/HHS N27220140045C, U19 AI100275 and AI118626, UM1 AI114271, P01 AI106695. Additional following grant(s) have supported as well this work: JHU OPP1109415, Umea University/EU Commission, U of Cape Town Gates Grant-OPP106626, and Emory U19AI111211.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <https://www.frontiersin.org/articles/10.3389/fimmu.2018.01369/full#supplementary-material>.

TABLE S1 | List of epitopes used as a positive dataset for the training set.

TABLE S2 | List of control peptides used as negative data for the training set.

TABLE S3 | Validation dataset description. (a) List of papers and corresponding number of peptide as positive, negative, and intermediate immunogenicity. (b) List of positive and negative peptides for the corresponding papers.

TABLE S4 | Additional training dataset from tetramer staining assays.

REFERENCES

- Peters B, Bui HH, Frankild S, Nielson M, Lundegaard C, Kostem E, et al. A community resource benchmarking predictions of peptide binding to MHC-I molecules. *PLoS Comput Biol* (2006) 2(6):e65. doi:10.1371/journal.pcbi.0020065
- Kumar N, Mohanty D. Structure-based identification of MHC binding peptides: benchmarking of prediction accuracy. *Mol Biosyst* (2010) 6(12):2508–20. doi:10.1039/c0mb00013b
- Hu X, Mamitsuka H, Zhu S. Ensemble approaches for improving HLA class I-peptide binding prediction. *J Immunol Methods* (2011) 374(1–2):47–52. doi:10.1016/j.jim.2010.09.007
- Trolle T, Metushi IG, Greenbaum JA, Kim Y, Sidney J, Lund O, et al. Automated benchmarking of peptide-MHC class I binding predictions. *Bioinformatics* (2015) 31(13):2174–81. doi:10.1093/bioinformatics/btv123
- Andreata M, Nielsen M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics* (2016) 32(4):511–7. doi:10.1093/bioinformatics/btv639
- Nielsen M, Andreata M. NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Med* (2016) 8(1):33. doi:10.1186/s13073-016-0288-x
- Jurtz V, Paul S, Andreata M, Marcatili P, Peters B, Nielsen M. NetMHCpan-4.0: improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J Immunol* (2017) 199(9):3360–8. doi:10.4049/jimmunol.1700893
- Karosiene E, Rasmussen M, Blicher T, Lund O, Buus S, Nielsen M. NetMHCIIpan-3.0, a common pan-specific MHC class II prediction method including all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ. *Immunogenetics* (2013) 65(10):711–24. doi:10.1007/s00251-013-0720-y
- Dhanda SK, Usmani SS, Agrawal P, Nagpal G, Gautam A, Raghava GPS. Novel in silico tools for designing peptide-based subunit vaccines and immunotherapeutics. *Brief Bioinform* (2017) 18(3):467–78. doi:10.1093/bib/bbw025
- Fleri W, Paul S, Dhanda SK, Mahajan S, Xu X, Peters B, et al. The immune epitope database and analysis resource in epitope discovery and synthetic

- vaccine design. *Front Immunol* (2017) 8:278. doi:10.3389/fimmu.2017.00278
11. Weiskopf D, Angelo MA, de Azeredo EL, Sidney J, Greenbaum JA, Fernando AN, et al. Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells. *Proc Natl Acad Sci U S A* (2013) 110(22):E2046–53. doi:10.1073/pnas.1305227110
 12. McKinney DM, Southwood S, Hinz D, Oseroff C, Arlehamn CS, Schulten V, et al. A strategy to determine HLA class II restriction broadly covering the DR, DP, and DQ allelic variants most commonly expressed in the general population. *Immunogenetics* (2013) 65(5):357–70. doi:10.1007/s00251-013-0684-y
 13. Greenbaum J, Sidney J, Chung J, Brander C, Peters B, Sette A. Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. *Immunogenetics* (2011) 63(6):325–35. doi:10.1007/s00251-011-0513-0
 14. Oseroff C, Sidney J, Kotturi MF, Kolla R, Alam R, Broide DH, et al. Molecular determinants of T cell epitope recognition to the common Timothy grass allergen. *J Immunol* (2010) 185(2):943–55. doi:10.4049/jimmunol.1000405
 15. Paul S, Lindestam Arlehamn CS, Scriba TJ, Dillon MB, Oseroff C, Hinz D, et al. Development and validation of a broad scheme for prediction of HLA class II restricted T cell epitopes. *J Immunol Methods* (2015) 422:28–34. doi:10.1016/j.jim.2015.03.022
 16. Yewdell JW, Bennink JR. Immunodominance in major histocompatibility complex class I-restricted T lymphocyte responses. *Annu Rev Immunol* (1999) 17:51–88. doi:10.1146/annurev.immunol.17.1.51
 17. Assarsson E, Sidney J, Oseroff C, Paschetto V, Bui HH, Frahm N, et al. A quantitative analysis of the variables affecting the repertoire of T cell specificities recognized after vaccinia virus infection. *J Immunol* (2007) 178(12):7890–901. doi:10.4049/jimmunol.178.12.7890
 18. Kotturi MF, Peters B, Buendia-Laysa F Jr, Sidney J, Oseroff C, Botten J, et al. The CD8+ T-cell response to lymphocytic choriomeningitis virus involves the L antigen: uncovering new tricks for an old virus. *J Virol* (2007) 81(10):4928–40. doi:10.1128/JVI.02632-06
 19. Stewart-Jones GB, McMichael AJ, Bell JI, Stuart DI, Jones EY. A structural basis for immunodominant human T cell receptor recognition. *Nat Immunol* (2003) 4(7):657–63. doi:10.1038/ni942
 20. Turner SJ, Doherty PC, McCluskey J, Rossjohn J. Structural determinants of T-cell receptor bias in immunity. *Nat Rev Immunol* (2006) 6(12):883–94. doi:10.1038/nri1977
 21. Kotturi MF, Scott I, Wolfe T, Peters B, Sidney J, Cheroutre H, et al. Naive precursor frequencies and MHC binding rather than the degree of epitope diversity shape CD8+ T cell immunodominance. *J Immunol* (2008) 181(3):2124–33. doi:10.4049/jimmunol.181.3.2124
 22. Jenkins MK, Chu HH, McLachlan JB, Moon JJ. On the composition of the preimmune repertoire of T cells specific for peptide-major histocompatibility complex ligands. *Annu Rev Immunol* (2010) 28:275–94. doi:10.1146/annurev-immunol-030409-101253
 23. Qi Q, Liu Y, Cheng Y, Glanville J, Zhang D, Lee JY, et al. Diversity and clonal selection in the human T-cell repertoire. *Proc Natl Acad Sci U S A* (2014) 111(36):13139–44. doi:10.1073/pnas.1409155111
 24. Frankild S, de Boer RJ, Lund O, Nielsen M, Kesmir C. Amino acid similarity accounts for T cell cross-reactivity and for “holes” in the T cell repertoire. *PLoS One* (2008) 3(3):e1831. doi:10.1371/journal.pone.0001831
 25. Calis JJ, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, et al. Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS Comput Biol* (2013) 9(10):e1003266. doi:10.1371/journal.pcbi.1003266
 26. Glanville J, Huang H, Nau A, Hatton O, Wagar LE, Rubelt F, et al. Identifying specificity groups in the T cell receptor repertoire. *Nature* (2017) 547(7661):94–8. doi:10.1038/nature22976
 27. Arlehamn CS, Sidney J, Henderson R, Greenbaum JA, James EA, Moutafsi M, et al. Dissecting mechanisms of immunodominance to the common tuberculosis antigens ESAT-6, CFP10, Rv2031c (hspX), Rv2654c (TB7.7), and Rv1038c (EsxJ). *J Immunol* (2012) 188(10):5020–31. doi:10.4049/jimmunol.1103556
 28. Lindestam Arlehamn CS, Gerasimova A, Mele F, Henderson R, Swann J, Greenbaum JA, et al. Memory T cells in latent *Mycobacterium tuberculosis* infection are directed against three antigenic islands and largely contained in a CXCR3+CCR6+ Th1 subset. *PLoS Pathog* (2013) 9(1):e1003130. doi:10.1371/journal.ppat.1003130
 29. Lindestam Arlehamn CS, McKinney DM, Carpenter C, Paul S, Rozot V, Makgohlho E, et al. A quantitative analysis of complexity of human pathogen-specific CD4 T cell responses in healthy *M. tuberculosis* infected South Africans. *PLoS Pathog* (2016) 12(7):e1005760. doi:10.1371/journal.ppat.1005760
 30. Schulten V, Greenbaum JA, Hauser M, McKinney DM, Sidney J, Kolla R, et al. Previously undescribed grass pollen antigens are the major inducers of T helper 2 cytokine-producing T cells in allergic individuals. *Proc Natl Acad Sci U S A* (2013) 110(9):3459–64. doi:10.1073/pnas.1300512110
 31. Westernberg L, Schulten V, Greenbaum JA, Natali S, Trippl V, McKinney DM, et al. T-cell epitope conservation across allergen species is a major determinant of immunogenicity. *J Allergy Clin Immunol* (2016) 138(2):571–8.e7. doi:10.1016/j.jaci.2015.11.034
 32. Hinz D, Oseroff C, Pham J, Sidney J, Peters B, Sette A. Definition of a pool of epitopes that recapitulates the T cell reactivity against major house dust mite allergens. *Clin Exp Allergy* (2015) 45(10):1601–12. doi:10.1111/cea.12507
 33. Dillon MB, Schulten V, Oseroff C, Paul S, Dullanty LM, Frazier A, et al. Different Blag T cell antigens dominate responses in asthma versus rhinitis subjects. *Clin Exp Allergy* (2015) 45(12):1856–67. doi:10.1111/cea.12643
 34. Weiskopf D, Bangs DJ, Sidney J, Kolla RV, De Silva AD, de Silva AM, et al. Dengue virus infection elicits highly polarized CX3CR1(+) cytotoxic CD4(+) T cells associated with protective immunity. *Proc Natl Acad Sci U S A* (2015) 112(31):E4256–63. doi:10.1073/pnas.1505956112
 35. Tangri S, Mothe BR, Eisenbraun J, Sidney J, Southwood S, Briggs K, et al. Rationally engineered therapeutic proteins with reduced immunogenicity. *J Immunol* (2005) 174(6):3187–96. doi:10.4049/jimmunol.174.6.3187
 36. Oseroff C, Pham J, Frazier A, Hinz D, Sidney J, Paul S, et al. Immunodominance in allergic T-cell reactivity to Japanese cedar in different geographic cohorts. *Ann Allergy Asthma Immunol* (2016) 117(6):680–689.e1. doi:10.1016/j.anai.2016.10.014
 37. Schulten V, Westernberg L, Birrueta G, Sidney J, Paul S, Busse P, et al. Allergen and epitope targets of mouse-specific T cell responses in allergy and asthma. *Front Immunol* (2018) 9:235. doi:10.3389/fimmu.2018.00235
 38. Oseroff C, Christensen LH, Westernberg L, Pham J, Lane J, Paul S, et al. Immunoproteomic analysis of house dust mite antigens reveals distinct classes of dominant T cell antigens according to function and serological reactivity. *Clin Exp Allergy* (2017) 47(4):577–92. doi:10.1111/cea.12829
 39. Bancroft T, Dillon MB, da Silva Antunes R, Paul S, Peters B, Crotty S, et al. Th1 versus Th2 T cell polarization by whole-cell and acellular childhood pertussis vaccines persists upon re-immunization in adolescence and adulthood. *Cell Immunol* (2016) 304–305:35–43. doi:10.1016/j.cellimm.2016.05.002
 40. Pham J, Oseroff C, Hinz D, Sidney J, Paul S, Greenbaum J, et al. Sequence conservation predicts T cell reactivity against ragweed allergens. *Clin Exp Allergy* (2016) 46(9):1194–205. doi:10.1111/cea.12772
 41. Antunes RDS, Paul S, Sidney J, Weiskopf D, Dan JM, Phillips E, et al. Definition of human epitopes recognized in tetanus toxoid and development of an assay strategy to detect ex vivo tetanus CD4(+) T cell responses. *PLoS One* (2017) 12(1):e0169086. doi:10.1371/journal.pone.0169086
 42. Manfredi AA, Protti MP, Wu XD, Howard JF Jr, Conti-Tronconi BM. CD4+ T-epitope repertoire on the human acetylcholine receptor alpha subunit in severe myasthenia gravis: a study with synthetic peptides. *Neurology* (1992) 42(5):1092–100. doi:10.1212/WNL.42.5.1092
 43. Herrera S, Escobar P, de Plata C, Avila GI, Corradin G, Herrera MA. Human recognition of T cell epitopes on the *Plasmodium vivax* circumsporozoite protein. *J Immunol* (1992) 148(12):3986–90.
 44. Sjöstedt A, Sandström G, Tarnvik A, Jaurin B. Nucleotide sequence and T cell epitopes of a membrane protein of *Francisella tularensis*. *J Immunol* (1990) 145(1):311–7.
 45. Rzepczyk CM, Csurhes PA, Baxter EP, Doran TJ, Irving DO, Kere N. Amino acid sequences recognized by T cells: studies on a merozoite surface antigen from the FCQ-27/PNG isolate of *Plasmodium falciparum*. *Immunol Lett* (1990) 25(1–3):155–63. doi:10.1016/0165-2478(90)90108-3
 46. Zevering Y, Houghten RA, Frazer IH, Good MF. Major population differences in T cell response to a malaria sporozoite vaccine candidate. *Int Immunol* (1990) 2(10):945–55. doi:10.1093/intimm/2.10.945
 47. Good MF, Pombo D, Quakyi IA, Riley EM, Houghten RA, Menon A, et al. Human T-cell recognition of the circumsporozoite protein of *Plasmodium falciparum*: immunodominant T-cell domains map to the polymorphic

- regions of the molecule. *Proc Natl Acad Sci U S A* (1988) 85(4):1199–203. doi:10.1073/pnas.85.4.1199
48. Carballido JM, Carballido-Perrig N, Kagi MK, Meloen RH, Wuthrich B, Heusser CH, et al. T cell epitope specificity in human allergic and nonallergic subjects to bee venom phospholipase A2. *J Immunol* (1993) 150(8 Pt 1):3582–91.
 49. Salvetti M, Ristori G, D'Amato M, Buttinelli C, Falcone M, Fieschi C, et al. Predominant and stable T cell responses to regions of myelin basic protein can be detected in individual patients with multiple sclerosis. *Eur J Immunol* (1993) 23(6):1232–9. doi:10.1002/eji.1830230606
 50. Bilsborough J, Carlisle M, Good MF. Identification of Caucasian CD4 T cell epitopes on the circumsporozoite protein of *Plasmodium vivax*. T cell memory. *J Immunol* (1993) 151(2):890–9.
 51. Manfredi AA, Protti MP, Dalton MW, Howard JF Jr, Conti-Tronconi BM. T helper cell recognition of muscle acetylcholine receptor in myasthenia gravis. Epitopes on the gamma and delta subunits. *J Clin Invest* (1993) 92(2):1055–67. doi:10.1172/JCI116610
 52. Moiola L, Protti MP, Manfredi AA, Yuen MH, Howard JF Jr, Conti-Tronconi BM. T-helper epitopes on human nicotinic acetylcholine receptor in myasthenia gravis. *Ann N Y Acad Sci* (1993) 681:198–218. doi:10.1111/j.1749-6632.1993.tb22887.x
 53. Atkinson MA, Bowman MA, Campbell L, Darrow BL, Kaufman DL, Maclaren NK. Cellular immunity to a determinant common to glutamate decarboxylase and coxsackie virus in insulin-dependent diabetes. *J Clin Invest* (1994) 94(5):2125–9. doi:10.1172/JCI117567
 54. Chaye H, Ou D, Chong P, Gillam S. Human T- and B-cell epitopes of E1 glycoprotein of rubella virus. *J Clin Immunol* (1993) 13(2):93–100. doi:10.1007/BF00919265
 55. Damhof RA, Drijfhout JW, Scheffer AJ, Wilterdink JB, Welling GW, Welling-Wester S. T cell responses to synthetic peptides of herpes simplex virus type 1 glycoprotein D in naturally infected individuals. *Arch Virol* (1993) 130(1–2):187–93. doi:10.1007/BF01319007
 56. Kellermann SA, McCormick DJ, Freeman SL, Morris JC, Conti-Fine BM. TSH receptor sequences recognized by CD4+ T cells in Graves' disease patients and healthy controls. *J Autoimmun* (1995) 8(5):685–98. doi:10.1006/jaut.1995.0051
 57. Muller CP, Ammerlaan W, Fleckenstein B, Krauss S, Kalbacher H, Schneider F, et al. Activation of T cells by the ragged tail of MHC class II-presented peptides of the measles virus fusion protein. *Int Immunol* (1996) 8(4):445–56. doi:10.1093/intimm/8.4.445
 58. Zhang L, Yang M, Chong P, Mohapatra SS. Multiple B- and T-cell epitopes on a major allergen of Kentucky Bluegrass pollen. *Immunology* (1996) 87(2):283–90. doi:10.1046/j.1365-2567.1996.467533.x
 59. Pender MP, Csurhes PA, Houghten RA, McCombe PA, Good MF. A study of human T-cell lines generated from multiple sclerosis patients and controls by stimulation with peptides of myelin basic protein. *J Neuroimmunol* (1996) 70(1):65–74. doi:10.1016/S0165-5728(96)00105-1
 60. Marttila J, Ilonen J, Lehtinen M, Parkkonen P, Salmi A. Definition of three minimal T helper cell epitopes of rubella virus E1 glycoprotein. *Clin Exp Immunol* (1996) 104(3):394–7. doi:10.1046/j.1365-2249.1996.54762.x
 61. Wang ZY, Okita DK, Howard J Jr, Conti-Fine BM. Th1 epitope repertoire on the alpha subunit of human muscle acetylcholine receptor in myasthenia gravis. *Neurology* (1997) 48(6):1643–53. doi:10.1212/WNL.48.6.1643
 62. Raulf-Heimsoth M, Chen Z, Rihs HP, Kalbacher H, Liebers V, Baur X. Analysis of T-cell reactive regions and HLA-DR4 binding motifs on the latex allergen Hev b 1 (rubber elongation factor). *Clin Exp Allergy* (1998) 28(3):339–48. doi:10.1046/j.1365-2222.1998.00230.x
 63. Kammerer R, Kettner A, Chvatchko Y, Dufour N, Tiercy JM, Corradin G, et al. Delineation of PLA2 epitopes using short or long overlapping synthetic peptides: interest for specific immunotherapy. *Clin Exp Allergy* (1997) 27(9):1016–26. doi:10.1111/j.1365-2222.1997.tb01253.x
 64. Flanagan KL, Plebanski M, Akinwunmi P, Lee EA, Reece WH, Robson KJ, et al. Broadly distributed T cell reactivity, with no immunodominant loci, to the pre-erythrocytic antigen thrombospondin-related adhesive protein of *Plasmodium falciparum* in West Africans. *Eur J Immunol* (1999) 29(6):1943–54. doi:10.1002/(SICI)1521-4141(199906)29:06<1943::AID-IMMU1943>3.0.CO;2-1
 65. Marttila J, Ilonen J, Norrby E, Salmi A. Characterization of T cell epitopes in measles virus nucleoprotein. *J Gen Virol* (1999) 80(Pt 7):1609–15. doi:10.1099/0022-1317-80-7-1609
 66. Lamonaca V, Missale G, Urbani S, Pilli M, Boni C, Mori C, et al. Conserved hepatitis C virus sequences are highly immunogenic for CD4(+) T cells: implications for vaccine development. *Hepatology* (1999) 30(4):1088–98. doi:10.1002/hep.510300435
 67. Woodfolk JA, Sung SS, Benjamin DC, Lee JK, Platts-Mills TA. Distinct human T cell repertoires mediate immediate and delayed-type hypersensitivity to the Trichophyton antigen, Tri r 2. *J Immunol* (2000) 165(8):4379–87. doi:10.4049/jimmunol.165.8.4379
 68. Stott LM, Barker RN, Urbaniak SJ. Identification of alloreactive T-cell epitopes on the Rhesus D protein. *Blood* (2000) 96(13):4011–9.
 69. Tejada-Simon MV, Hong J, Rivera VM, Zhang JZ. Reactivity pattern and cytokine profile of T cells primed by myelin peptides in multiple sclerosis and healthy individuals. *Eur J Immunol* (2001) 31(3):907–17. doi:10.1002/1521-4141(200103)31:3<907::AID-IMMU907>3.0.CO;2-1
 70. Marttila J, Juhela S, Vaarala O, Hyoty H, Roivainen M, Hinkkanen A, et al. Responses of coxsackievirus B4-specific T-cell lines to 2C protein-characterization of epitopes with special reference to the GAD65 homology region. *Virology* (2001) 284(1):131–41. doi:10.1006/viro.2001.0917
 71. Holen E, Bolann B, Elsayed S, Novel B and T cell epitopes of chicken ovomucoid (Gal d 1) induce T cell secretion of IL-6, IL-13, and IFN-gamma. *Clin Exp Allergy* (2001) 31(6):952–64. doi:10.1046/j.1365-2222.2001.01102.x
 72. Wertheimer AM, Miner C, Lewinsohn DM, Sasaki AW, Kaufman E, Rosen HR. Novel CD4+ and CD8+ T-cell determinants within the NS3 protein in subjects with spontaneously resolved HCV infection. *Hepatology* (2003) 37(3):577–89. doi:10.1053/jhep.2003.50115
 73. de Silva HD, Gardner LM, Drew AC, Beezhold DH, Rolland JM, O'Hehir RE. The hevein domain of the major latex-glove allergen Hev b 6.01 contains dominant T cell reactive sites. *Clin Exp Allergy* (2004) 34(4):611–8. doi:10.1111/j.1365-2222.2004.1919.x
 74. Elsayed S, Eriksen J, Oysaed LK, Idsoe R, Hill DJ. T cell recognition pattern of bovine milk alphaS1-casein and its peptides. *Mol Immunol* (2004) 41(12):1225–34. doi:10.1016/j.molimm.2004.05.010
 75. Sone T, Dairiki K, Morikubo K, Shimizu K, Tsunoo H, Mori T, et al. Identification of human T cell epitopes in Japanese cypress pollen allergen, Cha o 1, elucidates the intrinsic mechanism of cross-allergenicity between Cha o 1 and Cry j 1, the major allergen of Japanese cedar pollen, at the T cell level. *Clin Exp Allergy* (2005) 35(5):664–71. doi:10.1111/j.1365-2222.2005.02221.x
 76. Schulze zur Wiesch J, Lauer GM, Day CL, Kim AY, Ouchi K, Duncan JE, et al. Broad repertoire of the CD4+ Th cell response in spontaneously controlled hepatitis C virus infection includes dominant and highly promiscuous epitopes. *J Immunol* (2005) 175(6):3603–13. doi:10.4049/jimmunol.175.6.3603
 77. Sarobe P, Lasarte JJ, Garcia N, Civeira MP, Borrás-Cuesta F, Prieto J. Characterization of T-cell responses against immunodominant epitopes from hepatitis C virus E2 and NS4a proteins. *J Viral Hepat* (2006) 13(1):47–55. doi:10.1111/j.1365-2893.2005.00653.x
 78. Ruiter B, Tregoeat V, M'Rabet L, Garssen J, Bruijnzeel-Koomen CA, Knol EF, et al. Characterization of T cell epitopes in alphaS1-casein in cow's milk allergic, atopic and non-atopic children. *Clin Exp Allergy* (2006) 36(3):303–10. doi:10.1111/j.1365-2222.2006.02436.x
 79. Ma Y, Bogdanos DP, Hussain MJ, Underhill J, Bansal S, Longhi MS, et al. Polyclonal T-cell responses to cytochrome P450IID6 are associated with disease activity in autoimmune hepatitis type 2. *Gastroenterology* (2006) 130(3):868–82. doi:10.1053/j.gastro.2005.12.020
 80. Kasprzowicz V, Isa A, Tolfvenstam T, Jeffery K, Bowness P, Klennerman P. Tracking of peptide-specific CD4+ T-cell responses after an acute resolving viral infection: a study of parvovirus B19. *J Virol* (2006) 80(22):11209–17. doi:10.1128/JVI.01173-06
 81. Sukati H, Watson HG, Urbaniak SJ, Barker RN. Mapping helper T-cell epitopes on platelet membrane glycoprotein IIIa in chronic autoimmune thrombocytopenic purpura. *Blood* (2007) 109(10):4528–38. doi:10.1182/blood-2006-09-044388
 82. Schulze Zur Wiesch J, Lauer GM, Timm J, Kuntzen T, Neukamm M, Berical A, et al. Immunologic evidence for lack of heterologous protection following resolution of HCV in patients with non-genotype 1 infection. *Blood* (2007) 110(5):1559–69. doi:10.1182/blood-2007-01-069583

83. Immonen A, Kinnunen T, Sirven P, Taivainen A, Houitte D, Perasaari J, et al. The major horse allergen Equ c 1 contains one immunodominant region of T cell epitopes. *Clin Exp Allergy* (2007) 37(6):939–47. doi:10.1111/j.1365-2222.2007.02722.x
84. Malhotra I, Wamachi AN, Mungai PL, Mzungu E, Koech D, Muchiri E, et al. Fine specificity of neonatal lymphocytes to an abundant malaria blood-stage antigen: epitope mapping of *Plasmodium falciparum* MSP1(33). *J Immunol* (2008) 180(5):3383–90. doi:10.4049/jimmunol.180.5.3383
85. Masuyama K, Chikamatsu K, Ikagawa S, Matsuoka T, Takahashi G, Yamamoto T, et al. Analysis of helper T cell responses to Cry j 1-derived peptides in patients with nasal allergy: candidate for peptide-based immunotherapy of Japanese cedar pollinosis. *Allergol Int* (2009) 58(1):63–70. doi:10.2332/allergolint.08-OA-0008
86. Sone T, Daijiki K, Morikubo K, Shimizu K, Tsunoo H, Mori T, et al. Recognition of T cell epitopes unique to Cha o 2, the major allergen in Japanese cypress pollen, in allergic patients cross-reactive to Japanese cedar and Japanese cypress pollen. *Allergol Int* (2009) 58(2):237–45. doi:10.2332/allergolint.08-OA-0027
87. Madsen D, Cantwell ER, O'Brien T, Johnson PA, Mahon BP. Adeno-associated virus serotype 2 induces cell-mediated immune responses directed against multiple epitopes of the capsid protein VP1. *J Gen Virol* (2009) 90 (Pt 11):2622–33. doi:10.1099/vir.0.014175-0
88. Pastorello EA, Monza M, Pravettoni V, Longhi R, Bonara P, Scibilia J, et al. Characterization of the T-cell epitopes of the major peach allergen Pru p 3. *Int Arch Allergy Immunol* (2010) 153(1):1–12. doi:10.1159/000301573
89. Matsuya N, Komori M, Nomura K, Nakane S, Fukudome T, Goto H, et al. Increased T-cell immunity against aquaporin-4 and proteolipid protein in neuromyelitis optica. *Int Immunol* (2011) 23(9):565–73. doi:10.1093/intimm/dxr056
90. Chaduvula M, Murtaza A, Misra N, Narayan NP, Ramesh V, Prasad HK, et al. Lsr2 peptides of *Mycobacterium leprae* show hierarchical responses in lymphoproliferative assays, with selective recognition by patients with anergic lepromatous leprosy. *Infect Immun* (2012) 80(2):742–52. doi:10.1128/IAI.05384-11
91. Etto T, de Boer C, Prickett S, Gardner LM, Voskamp A, Davies JM, et al. Unique and cross-reactive T cell epitope peptides of the major Bahia grass pollen allergen, Pas n 1. *Int Arch Allergy Immunol* (2012) 159(4):355–66. doi:10.1159/000338290
92. Ravkov EV, Pavlov IY, Martins TB, Gleich GJ, Wagner LA, Hill HR, et al. Identification and validation of shrimp-tropomyosin specific CD4 T cell epitopes. *Hum Immunol* (2013) 74(12):1542–9. doi:10.1016/j.humimm.2013.08.276
93. Schwaiger J, Aberle JH, Stiasny K, Knapp B, Schreiner W, Fae I, et al. Specificities of human CD4+ T cell responses to an inactivated flavivirus vaccine and infection: correlation with structure and epitope prediction. *J Virol* (2014) 88(14):7828–42. doi:10.1128/JVI.00196-14
94. Ronka AL, Kinnunen TT, Goudet A, Rytönen-Nissinen MA, Sairanen J, Kailaanmaki AH, et al. Characterization of human memory CD4(+) T-cell responses to the dog allergen Can f 4. *J Allergy Clin Immunol* (2015) 136(4):1047–54.e10. doi:10.1016/j.jaci.2015.02.025
95. Kailaanmaki A, Kinnunen T, Ronka A, Rytönen-Nissinen M, Lidholm J, Mattsson L, et al. Human memory CD4+ T cell response to the major dog allergen Can f 5, prostatic kallikrein. *Clin Exp Allergy* (2016) 46(5):720–9. doi:10.1111/cea.12694
96. Oshima M, Deitiker P, Jankovic J, Aoki KR, Atassi MZ. Submolecular recognition of the C-terminal domain of the heavy chain of botulinum neurotoxin type A by T cells from toxin-treated cervical dystonia patients. *Immunobiology* (2016) 221(4):568–76. doi:10.1016/j.imbio.2015.12.002
97. Gaido CM, Stone S, Chopra A, Thomas WR, Le Souef PN, Hales BJ. Immunodominant T-cell epitopes in the VP1 capsid protein of rhinovirus species A and C. *J Virol* (2016) 90(23):10459–71. doi:10.1128/JVI.01701-16
98. Oshima M, Deitiker P, Jankovic J, Atassi MZ. Submolecular recognition regions of the HN domain of the heavy chain of botulinum neurotoxin type A by T cells from toxin-treated cervical dystonia patients. *J Neuroimmunol* (2016) 300:36–46. doi:10.1016/j.jneuroim.2016.09.013
99. Paul S, Sidney J, Sette A, Peters B. TepiTool: a pipeline for computational prediction of T cell epitope candidates. *Curr Protoc Immunol* (2016) 114:18.19.1–18.19.24. doi:10.1002/cpim.12
100. Weiskopf D, Angelo MA, Grifoni A, O'Rourke PH, Sidney J, Paul S, et al. HLA-DRB1 alleles are associated with different magnitudes of dengue virus-specific CD4+ T-cell responses. *J Infect Dis* (2016) 214(7):1117–24. doi:10.1093/infdis/jiw309
101. Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, et al. The immune epitope database (IEDB) 3.0. *Nucleic Acids Res* (2015) 43(Database issue):D405–12. doi:10.1093/nar/gku938
102. Andreatta M, Schafer-Nielsen C, Lund O, Buus S, Nielsen M. NNAlign: a web-based prediction method allowing non-expert end-user discovery of sequence motifs in quantitative peptide data. *PLoS One* (2011) 6(11):e26781. doi:10.1371/journal.pone.0026781
103. Swets JA. Measuring the accuracy of diagnostic systems. *Science* (1988) 240(4857):1285–93. doi:10.1126/science.3287615
104. Kim Y, Ponomarenko J, Zhu Z, Tamang D, Wang P, Greenbaum J, et al. Immune epitope database analysis resource. *Nucleic Acids Res* (2012) 40(Web Server issue):W525–30. doi:10.1093/nar/gks438
105. Vacic V, Iakoucheva LM, Radivojac P. Two sample logo: a graphical representation of the differences between two sets of sequence alignments. *Bioinformatics* (2006) 22(12):1536–7. doi:10.1093/bioinformatics/btl151
106. Dhanda SK, Gupta S, Vir P, Raghava GP. Prediction of IL4 inducing peptides. *Clin Dev Immunol* (2013) 2013:263952. doi:10.1155/2013/263952
107. Dhanda SK, Vir P, Raghava GP. Designing of interferon-gamma inducing MHC class-II binders. *Biol Direct* (2013) 8:30. doi:10.1186/1745-6150-8-30
108. Nagpal G, Usmani SS, Dhanda SK, Kaur H, Singh S, Sharma M, et al. Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential. *Sci Rep* (2017) 7:42851. doi:10.1038/srep42851
109. Sant'Angelo DB, Robinson E, Janeway CA Jr, Denzin LK. Recognition of core and flanking amino acids of MHC class II-bound peptides by the T cell receptor. *Eur J Immunol* (2002) 32(9):2510–20. doi:10.1002/1521-4141(200209)32:9<2510::AID-IMMU2510>3.0.CO;2-Q
110. Nielsen M, Lundegaard C, Lund O, Kesmir C. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* (2005) 57(1–2):33–41. doi:10.1007/s00251-005-0781-7
111. Yin L, Calvo-Calle JM, Dominguez-Amoroso O, Stern LJ. HLA-DM constrains epitope selection in the human CD4 T cell response to vaccinia virus by favoring the presentation of peptides with longer HLA-DM-mediated half-lives. *J Immunol* (2012) 189(8):3983–94. doi:10.4049/jimmunol.1200626
112. Alexander J, Sidney J, Southwood S, Ruppert J, Oseroff C, Maewal A, et al. Development of hi0067h potency universal DR-restricted helper epitopes by modification of high affinity DR-blocking peptides. *Immunity* (1994) 1(9):751–61. doi:10.1016/S1074-7613(94)80017-0
113. Hung CF, Tsai YC, He L, Wu TC. DNA vaccines encoding Ii-PADRE generates potent PADRE-specific CD4+ T-cell immune responses and enhances vaccine potency. *Mol Ther* (2007) 15(6):1211–9. doi:10.1038/sj.mt.6300121
114. Carrasco Pro S, Sidney J, Paul S, Lindestam Arlehamn C, Weiskopf D, Peters B, et al. Automatic generation of validated specific epitope sets. *J Immunol Res* (2015) 2015:763461. doi:10.1155/2015/763461
115. Weiskopf D, Cerpas C, Angelo MA, Bangs DJ, Sidney J, Paul S, et al. Human CD8+ T-cell responses against the 4 dengue virus serotypes are associated with distinct patterns of protein targets. *J Infect Dis* (2015) 212(11):1743–51. doi:10.1093/infdis/jiv289

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling Editor declared a past co-authorship with the authors.

Copyright © 2018 Dhanda, Karosiene, Edwards, Grifoni, Paul, Andreatta, Weiskopf, Sidney, Nielsen, Peters and Sette. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.